UNIVERSITI TUNKU ABDUL RAHMAN

Department of Mathematics and Actuarial Science

CONTENTS

3 Normal Theory Inference

3.1 Normal Distribution

Definition 1.
A random variable Y with density function

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

is said to have a **normal** (*Gaussian*) **distribution** with

$$E(Y) = \mu \quad \text{and} \quad V(Y) = \sigma^2.$$

We will use the notation

$$Y \sim N(\mu, \sigma^2)$$

Suppose Z has a normal distribution with $E(Z) = 0$ and $V(Z) = 1$, i.e.,

$$Z \sim N(0, 1),$$

then Z is said to have a *standard normal* distribution.

Definition 2.
Suppose $\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$ is a random vector whose elements are independently distributed standard normal random variables. For any $m \times n$ matrix A, We say that

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}^T\mathbf{Z}$$

has a ***multivariate normal distribution*** with mean vector

$$\begin{aligned} E(\mathbf{y}) &= E(\boldsymbol{\mu} + \mathbf{A^T Z}) \\ &= \boldsymbol{\mu} + \mathbf{A^T} E(\mathbf{Z}) \\ &= \boldsymbol{\mu} + \mathbf{A^T 0} \\ &= \boldsymbol{\mu} \end{aligned}$$

and variance-covariance matrix

$$\begin{aligned} V(\mathbf{y}) &= \mathbf{A^T} V(\mathbf{Z})\mathbf{A} \\ &= \mathbf{A^T A} \equiv \boldsymbol{\Sigma} \end{aligned}$$

We will use the notation

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

When $\boldsymbol{\Sigma}$ is positive definite, the joint density function is

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\, e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{y}-\boldsymbol{\mu})}$$

The multivariate normal distribution has many useful properties:

Result 1. Normality is preserved under linear transformations: If

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

then

$$w = \mathbf{c^T y} \sim N(\mathbf{c}^T\boldsymbol{\mu}, \mathbf{c}^T\boldsymbol{\Sigma}\mathbf{c})$$
$$\mathbf{W} = \mathbf{c} + B\mathbf{y} \sim N(\mathbf{c} + B\boldsymbol{\mu}, B\boldsymbol{\Sigma}B^T)$$

for any non-random $\mathbf{c}$ and B.

Result 2.
Suppose

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \right)$$

then

$$\mathbf{y}_1 \sim N(\boldsymbol{\mu}_1, \mathbf{\Sigma}_{11}).$$

Note: This result applies to any subset of the elements of $\mathbf{y}$ because you can move that subset to the top of the vector by multiplying $\mathbf{y}$ by an appropriate matrix of zeros and ones.

Example 1. Suppose

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 1 & -1 \\ 1 & 3 & -3 \\ -1 & -3 & 9 \end{bmatrix} \right)$$

Find the distribution of

(a) y_1

(b) y_2

(c) y_3

(d) $\begin{bmatrix} y_1 \\ y_3 \end{bmatrix}$

If $w_1 = y_1 - 2y_2 + y_3$ and $w_2 = 3y_1 + y_2 - 2y_3$, then find the distribution of

(e) w_1

(f) w_2

(g) $\mathbf{W} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

Comment:
If $\mathbf{y}_1 \sim N(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ and $\mathbf{y}_2 \sim N(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$, it is **not** always true that $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$ has a normal distribution.

Result 3.
If $\mathbf{y}_1$ and $\mathbf{y}_2$ are **independent** random vectors such that

$$\mathbf{y}_1 \sim N(\boldsymbol{\mu}_1, \mathbf{\Sigma_1}) \quad \text{and} \quad \mathbf{y}_2 \sim N(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$$

then

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \mathbf{\Sigma}_2 \end{bmatrix} \right)$$

Result 4.
If $\mathbf{y}^T = [\mathbf{y}_1 \cdots \mathbf{y}_k]$ is a random vector with a multivariate normal distribution, then $\mathbf{y}_1$, $\mathbf{y}_2$, $\cdots$, $\mathbf{y}_k$ are **independent** if and only if $Cov(\mathbf{y}_i, \mathbf{y}_j) = 0$ for all $i \neq j$.

Comments:

(i) If $\mathbf{y}_i$ is independent of $\mathbf{y}_j$, then $Cov(\mathbf{y}_i, \mathbf{y}_j) = 0$.

(ii) When $\mathbf{y} = (y_1, \cdots, y_n)^T$ has a multivariate normal distribution, y_i uncorrelated with y_j implies y_i is independent of y_j. This is usually not true for other distributions.

Result 5.
If
$$\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_y\\ \boldsymbol{\mu}_x\end{bmatrix}\begin{bmatrix}\Sigma_{yy} & \Sigma_{yx}\\ \Sigma_{xy} & \Sigma_{xx}\end{bmatrix}\right)$$
with a positive definite covariance matrix, the **conditional distribution** of $\mathbf{y}$ given the value of $\mathbf{X}$ is a normal distribution with mean vector

$$E(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu}_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$
and positive definite covaraince matrix
$$V(\mathbf{y}|\mathbf{x}) = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$$

note that this does not depend on the value of $\mathbf{x}$

3.2 Quadratic forms: $\mathbf{y^TAy}$

Some useful information about the distribution of quadratic forms is summarized in the following results.

Result 6.
If $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ is a random vector with
$$E(\mathbf{y}) = \boldsymbol{\mu}$$
and
$$V(\mathbf{y}) = \boldsymbol{\Sigma}$$
and $\mathbf{A}$ is an $n \times n$ non-random matrix, then
$$E(\mathbf{y^T A y}) = \boldsymbol{\mu}^T \mathbf{A}\boldsymbol{\mu} + tr(\mathbf{A}\boldsymbol{\Sigma})$$

..

Example 2.
Consider a Gauss-Markov model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } V(\mathbf{y}) = \sigma^2 I.$$

Show that $\hat{\sigma}^2 = \frac{SSE}{n - rank(\mathbf{X})}$ is an unbiased estimator of σ^2.

..

3.3 Chi-square Distributions

Definition 3.

Let $\mathbf{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim N(\mathbf{0}, I)$, i.e., the elements of Z are n independent standard normal random variables. The distribution of

$$W = \mathbf{Z}^T\mathbf{Z} = \sum_{i=1}^{n} Z_i^2$$

is called the **central chi-square distribution** with n degrees of freedom.

We will use the notation

$$W \sim \boldsymbol{\chi}^2_{(n)}$$

The density function is

$$f(w) = \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} w^{n/2-1} e^{-w/2}$$

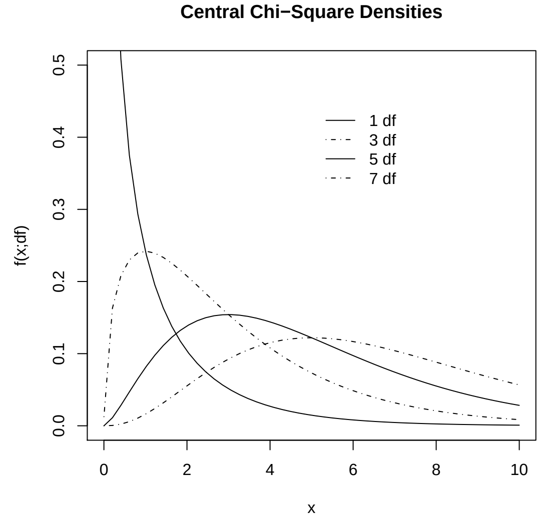
Moments:
If $W \sim \boldsymbol{\chi}^2_n$, then

(i) $E(W) = n$

(ii) $V(W) = 2n$

(iii) $M_W(t) = E(e^{tW}) = \frac{1}{(1-2t)^{n/2}}$

Note: The R-codes is store in the file: chidenR.txt.



Definition 4.
Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, I)$$

i.e., the elements of $\mathbf{y}$ are independent normal random variables with $y_i \sim N(\mu_i, 1)$. The distribution of the random variable

$$W = \mathbf{y}^T\mathbf{y} = \sum_{i=1}^{n} y_i^2$$

is called a **noncentral chi-square distribution** with n degrees of freedom and noncentrality parameter

$$\lambda \;=\; \boldsymbol{\mu}^T\boldsymbol{\mu} \;=\; \sum_{i=1}^{n} \mu_i^2$$

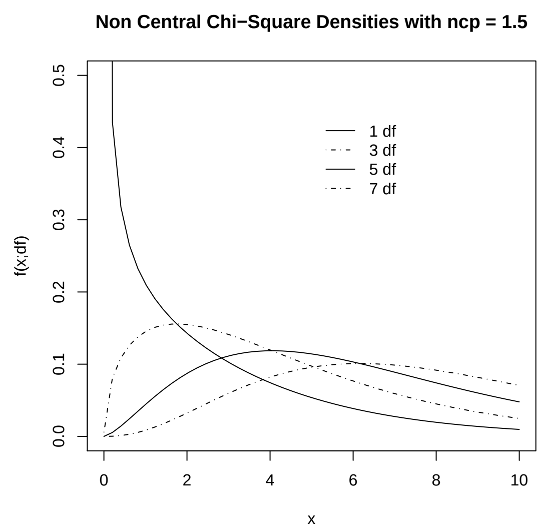
We will use the notation

$$W \sim \boldsymbol{\chi}_n^2(\lambda)$$

The density function is:

$$f(w) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \frac{w^{\frac{1}{2}n+k-1} e^{-w/2}}{2^{\frac{1}{2}n+k}\Gamma(\frac{1}{2}n+k)}$$

Note: The R-codes is store in the file: ncchidenR.txt.



3.4 F Distribution

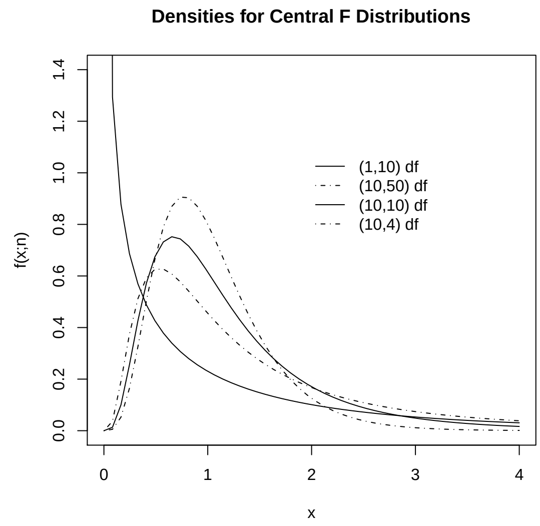
Definition 5.
If $W_1 \sim \boldsymbol{\chi}_{n_1}^2$ and $W_2 \sim \boldsymbol{\chi}_{n_2}^2$ and W_1 and W_2 are **independent**, then the distribution of

$$F = \frac{W_1/n_1}{W_2/n_2}$$

is called the **central F distribution** with n_1 and n_2 degrees of freedom.
We will use the notation

$$F \sim F_{n_1,n_2}$$

Note: The R-codes is store in the file: fdenR.txt.



Definition 6.
If $W_1 \sim \boldsymbol{\chi}^2_{n_1}(\lambda)$ and $W_2 \sim \boldsymbol{\chi}^2_{n_2}$ and W_1 and W_2 are **independent**, then the distribution of

$$F = \frac{W_1/n_1}{W_2/n_2}$$

is called a **noncentral F distribution** with n_1 and n_2 degrees of freedom and noncentrality parameter λ.

We will use the notation

$$F \sim F_{n_1,n_2}(\lambda)$$

3.5 Students's $t-$distribution

Definition 7.
If $Z \sim N(0,1)$ and $W \sim \chi^2_n$and Z and$W$ are independent, then the distribution of

$$T = \frac{Z}{\sqrt{W/n}}$$

is called a student's $t-$distribution with n degrees of freedom.

Its density function is

$$f(t) = \frac{\Gamma(\frac{1}{2}n + \frac{1}{2})}{\sqrt{n\pi}\Gamma(\frac{1}{2}n)} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}$$

We will use the notation

$$T \sim t_n$$

Definition 8.
If $y \sim N(\mu, 1)$ and $W \sim \chi^2_n$ and y and$W$ are independent, then the distribution of

$$T = \frac{Z}{W/n}$$

is called a noncentral student's $t-$distribution with n degrees of freedom and non-central parameter μ.

We will use the notation

$$T \sim t_n(\mu)$$

The density function is:

$$f(t) = \frac{n^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \frac{e^{-\frac{1}{2}\mu^2}}{(n+t^2)^{\frac{1}{2}(n+1)}} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{1}{2}n + \frac{1}{2}k + \frac{1}{2})\mu^k 2^{\frac{1}{2}k} t^k}{k!(n+t^2)^{\frac{1}{2}k}}$$

3.6 Sums of squares in ANOVA tables

Sums of squares in ANOVA tables are quadratic forms

$$\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y}$$

where $\mathbf{A}$ is a non-negative definite symmetric matrix (**usually a projection matrix**).

To develop F-tests we need to identify conditions under which

- $\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y}$ has a central (or noncentral) chi-square distribution
- $\mathbf{y}^{\mathbf{T}}\mathbf{A}_{\mathbf{i}}\mathbf{y}$ and $\mathbf{y}^{\mathbf{T}}\mathbf{A}_{\mathbf{j}}\mathbf{y}$ are independent

Result 7.
Let $\mathbf{A}$ be an $n \times n$ symmetric matrix with rank$(\mathbf{A}) = k$, and let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive definite matrix. If

$$\mathbf{A}\boldsymbol{\Sigma} \text{ is idempotent}$$

then

$$\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y} \sim \chi^2_k\left(\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}\right)$$

In addition, if $\mathbf{A}\boldsymbol{\mu} = \mathbf{0}$ then

$$\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y} \sim \chi^2_k$$

..

..

Example 3.
For the Gauss-Markov model with

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{and} \quad V(\mathbf{y}) = \sigma^2\mathbf{I}$$

include the assumption that

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \sim N(X\boldsymbol{\beta}, \sigma^2\mathbf{I}).$$

Show that $\frac{SSE}{\sigma^2} \sim \chi^2_{n-k}$.

Example 4.

Continuing Example 3, show that $\frac{1}{\sigma^2}\sum_{i=1}^{n} \hat{\mathbf{y}}_i^2 \sim \chi^2(\lambda)$, where λ is the non-central parameter.

The next result addresses the independence of several quadratic forms

Result 8.
Let $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

and let $\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A_p}$ be $n \times n$ symmetric matrices. If

$$\mathbf{A_i}\boldsymbol{\Sigma}\mathbf{A_j} = 0 \text{ for all } i \neq j$$

then

$$\mathbf{y^T A_1 y},\ \mathbf{y^T A_2 y},\ \ldots,\ \mathbf{y^T A_p y}$$

are independent random variables.

Example 5.
Continuing Example 3, show that the "uncorrected" model sum of squares

$$\sum_{i=1}^{n} \hat{y}_i^2 = \mathbf{y^T P_X y}$$

and the sum of squared residuals

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \mathbf{y^T (I - P_X) y}$$

are independently distributed for the "normal theory" Gauss-Markov model where

$$\mathbf{y} \sim N(X\boldsymbol{\beta}, \sigma^2\mathbf{I}).$$

Example 6.
If $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T \sim N(\mu\mathbf{1}, \sigma^2\mathbf{I})$. Find the distribution of $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sigma^2}$.

Example 7.
Suppose that $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

Let

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & -8 \\ -3 & 2 & -6 \\ -8 & -6 & 3 \end{bmatrix}$$

(a) Does $\mathbf{y^TAy}$ have a chi-square distribution?

(b) If $\boldsymbol{\Sigma} = \sigma^2 I$, does $\mathbf{y^TAy}/\sigma^2$ have a chi-square distribution?

Example 8. Suppose that $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix}, \mathbf{A} = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix}.$$

(a) What is the distribution of $\mathbf{y^TAy}/\sigma^2$?

(b) Are $\mathbf{y^TAy}$ and $\mathbf{By}$ independent?

(c) Are $\mathbf{y^TAy}$ and $y_1 + y_2 + y_3$ independent?

Example 9.

Consider the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$
$$i = 1, 2; j = 1, 2, 3; k = 1, 2$$

where $\epsilon_{ijk} \sim NID(0, \sigma^2)$, α_i is associated with the i-th level of factor A, β_j is associated with the j-th level of factor B, and γ_{ij} is an interaction parameter.

(a) Define $SSE = \sum_{i1=}^{2} \sum_{j=1}^{3} \sum_{k=1}^{2} (y_{ijk} - \bar{y}_{ij\bullet})^2$, where $\bar{y}_{ij\bullet} = \frac{1}{2}(y_{ij1} + y_{ij2})$. Show that $\frac{SSE}{\sigma^2}$ has a chi-squares distribution. States the degrees of fredom.

(b) Consider the estimator

$$\hat{C} = \bar{y}_{\bullet 3 \bullet} - \bar{y}_{\bullet 1 \bullet},$$

where

$$\bar{y}_{\bullet j \bullet} = \frac{1}{4} \sum_{i=1}^{2} \sum_{k=1}^{2} y_{ijk}.$$

Show that

$$F = \frac{m(\hat{C})^2}{SSE}$$

has an F-distribution for some constant m. Report the value of m and the degrees of freedom for the F-distribution.

3.7 Hypotesis Test for $E(\mathbf{y})$

In Example 3 we showed that

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} \hat{y}_i^2 \sim \chi_k^2 \left(\frac{\boldsymbol{\beta}^T \mathbf{X}^\mathbf{T} \mathbf{X} \boldsymbol{\beta}}{2\sigma^2} \right)$$

and

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \sim \chi_{n-k}^2$$

where $k = \text{rank}(\mathbf{X})$.

By Defn 6,

$$F = \frac{\frac{1}{k\sigma^2} \sum_{i=1}^{n} \hat{y}_i^2}{\frac{1}{(n-k)\sigma^2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

uncorrected model

$\downarrow$ mean square

$$= \frac{\frac{1}{k} \sum_{i=1}^{n} \hat{y}_i^2}{\frac{1}{n-k} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

$\nearrow$ Residual mean square

$$\sim F_{k,n-k} \left(\tfrac{1}{2\sigma^2} \boldsymbol{\beta}^T \mathbf{X}^\mathbf{T} \mathbf{X} \boldsymbol{\beta} \right)$$

$\uparrow$

This reduces to a central F distribution with $(k, n-k)$ d.f. when $\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$

Use

$$F = \frac{\frac{1}{k} \sum_{i=1}^{n} \hat{y}_i^2}{\frac{1}{n-k} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

to test the null hypothesis

$$H_0 : E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

against the alternative

$$H_A : E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \neq \mathbf{0}$$

Comments

(i) The null hypothesis corresponds to the condition under which F has a central F distribution (**the non-centrality parameter is zero**).

$$\lambda = \frac{1}{2\sigma^2} (\mathbf{X}\boldsymbol{\beta})^T (\mathbf{X}\boldsymbol{\beta}) = 0$$

if and only if $\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$.

(ii) If $k = \text{rank}(\mathbf{X}) =$ number of columns in $\mathbf{X}$, then $H_0 : \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$ is equivalent to $H_0 : \boldsymbol{\beta} = \mathbf{0}$.

(iii) If $k = \text{rank}(X)$ is less than the number of columns in $\mathbf{X}$, then $\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$ for some $\boldsymbol{\beta} \neq \mathbf{0}$ and $H_0 : \mathbf{X}\boldsymbol{\beta} = 0$ is **not** equivalent to $H_0 : \boldsymbol{\beta} = \mathbf{0}$.

Example 4 is a simple illustration of a typical

$$\begin{aligned}
\sum_{i=1}^{n} y_i^2 &= \mathbf{y}^T\mathbf{y} \\
&= \mathbf{y}^T\left[(\mathbf{I} - \mathbf{P_X}) + \mathbf{P_X}\right]\mathbf{y} \\
&= \mathbf{y}^T\underbrace{(\mathbf{I} - \mathbf{P_X})}_{\text{call this } \mathbf{A_2}}\mathbf{y} + \mathbf{y^T}\underbrace{\mathbf{P_X}}_{\text{call this } \mathbf{A_1}}\mathbf{y} \\
&= \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{d.f.} = \text{rank}(\mathbf{A_2})} + \underbrace{\sum_{i=1}^{n}\hat{y}_i^2}_{\text{d.f.} = \text{rank}(\mathbf{A_1})}
\end{aligned}$$

More generally an uncorrected total sum of squares can be partitioned as

$$\begin{aligned}
\sum_{i=1}^{n} y_i^2 &= \mathbf{y^T y} \\
&= \mathbf{y^T A_1 y} + \mathbf{y^T A_2 y} + \\
&= \cdots + \mathbf{y^T A_k y}
\end{aligned}$$

using orthogonal projection matrices

$$\mathbf{A_1} + \mathbf{A_2} + \cdots + \mathbf{A_k} = \mathbf{I_{n\times n}}$$

where

$$\text{rank}(\mathbf{A_1}) + \text{rank}(\mathbf{A_2}) + \cdots + \text{rank}(\mathbf{A_k}) = n$$

and

$$\mathbf{A_i A_j} = \mathbf{0} \quad \text{for any } i \neq j\,.$$

Since we are dealing with orthogonal projection matrices we also have

$$\mathbf{A_i^T} = \mathbf{A_i} \qquad \text{(symmetry)}$$

$$\mathbf{A_i A_i} = \mathbf{A_i} \qquad \text{(idempodent matrices)}$$

Result 9.
Let $\mathbf{A_1}, \mathbf{A_2}, \cdots, \mathbf{A_k}$ be $n \times n$ symmetric matrices such that

$$\mathbf{A_1} + \mathbf{A_2} + \cdots + \mathbf{A_k} = \mathbf{I}\,.$$

Then the following statments are equivalent

(i) $\mathbf{A_i A_j} = \mathbf{0}$ for any $i \neq j$

(ii) $\mathbf{A_i A_i} = \mathbf{A_i}$ for all $i = 1, \ldots, k$

(iii) $\text{rank}(\mathbf{A_1}) + \cdots + \text{rank}(\mathbf{A_k}) = n$

..

..

Result 10. (Cochran's Theorem)

Let $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, \sigma^2 I)$

and let $\mathbf{A_1}, \mathbf{A_2}, \cdots, \mathbf{A_k}$ be $n \times n$ symmetric matrices with

$$\mathbf{I} = \mathbf{A_1} + \mathbf{A_2} + \cdots + \mathbf{A_k}$$

and

$$n = r_1 + r_2 + \cdots + r_k$$

where $r_i = \text{rank}(\mathbf{A_i})$. Then, for $i = 1, 2, \ldots, k$

$$\frac{1}{\sigma^2}\mathbf{y^T A_i y} \sim \chi^2_{r_i}\left(\frac{1}{\sigma^2}\boldsymbol{\mu}^{\mathbf{T}}\mathbf{A_i}\boldsymbol{\mu}\right)$$

and

$$\mathbf{y^T A_1 y},\ \mathbf{y^T A_2 y},\ \cdots,\ \mathbf{y^T A_k y}$$

are distributed independently.

Example 10.

When gasoline is pumped into the tank of a car, vapors are vented into the atmosphere. A company has developed a device that can be installed in the gas tank of a car to prevent vapors from escaping when the gas tank is filled. A small study was performed to examine the effectiveness of this device. Four cars were used in the study, and the device was installed in the gas tank of two of the cars. Gasoline was pumped into the tank of each car and the amount of gas vapor that escaped (y) was measured. Since the temperature of the gasoline (X_1) can affect the outcome, two gasoline temperatures were used. In this study, $X_2 = 1$ if the device was used and $X_2 = -1$ if the device was not installed in the gas tank.

Amount of vapor that escapes Y	Gasoline temperature (° C) X_1	Use of device X_2
y_1	0	-1
y_2	30	-1
y_3	0	1
y_4	30	1

Consider the model

$$y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

where ϵ_i is independently and identically distributed $N(0, \sigma^2)$.

(a) To test for a device effect, the researchers propose the following test statistic

$$F = \frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{2SSE}$$

Show that this statistic has an F-distribution. Report its degrees of freedom.

(b) With respect to $\boldsymbol{\beta} = (\beta_1 \quad \beta_2)^T$, describe the null hypothesis that can be tested with the F-test in Part (a). What is the alternative hypothesis?

Example 11.

Suppose $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Define $\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_{16} \end{bmatrix}$ and $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ and $\mathbf{P_x} = \mathbf{X}(\mathbf{X^T X})^{-}\mathbf{X^T}$ and $\mathbf{P_1} = \mathbf{1}(\mathbf{1^T 1})^{-}\mathbf{1^T}$. Apply Cochran's Theorem to find the distribution of $\frac{1}{\sigma^2}\mathbf{Y^T}(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y}$ and $\frac{1}{\sigma^2}\mathbf{Y^T}(\mathbf{I} - \mathbf{P_X})\mathbf{Y}$. Then, derive the distribution of $V = \dfrac{c\mathbf{Y^T}(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y}}{\mathbf{Y^T}(\mathbf{I} - \mathbf{P_X})\mathbf{Y}}$. Report c, degrees of freedom and a formula for the noncentrality parameter.

Example 12.

Suppose $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Define

$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_{16} \end{bmatrix}$; $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$; $\mathbf{P_x} = \mathbf{X}(\mathbf{X^T X})^{-}\mathbf{X^T}$ and $\mathbf{P_1} = \mathbf{1}(\mathbf{1^T 1})$

Show that

$$\frac{1}{\sigma^2}(\mathbf{X}\boldsymbol{\beta})^T(\mathbf{P_X} - \mathbf{P_1})\mathbf{X}\boldsymbol{\beta} = \frac{\beta_1^2 \sum_{i=1}^n (X_i - \bar{X})^2)}{\sigma^2}$$